Anirudh Madhigiri Gopinath

Munich, Germany
anirudh.athreya1@gmail.com ● [+4917685926258]
https://www.linkedin.com/in/anirudhmg/ ● https://github.com/AnirudhMadhigiriGopinath

Summary

As a Data Scientist with 3+ years of experience, I have developed a strong foundation in machine learning, deep learning, and data engineering. I specialize in creating scalable models, building end-to-end data pipelines, and deploying AI solutions on AWS. My record includes increasing the efficiency of the workflow by 40% and achieving 90% accuracy in NLP-based classification tasks. My technical skills, combined with a problem-solving mindset, allow me to deliver impactful data solutions in fast-paced environments. I hold a Master's degree specialized in Data Science. I speak B1-level German and am very excited to learn B2. I am willing to relocate and start immediately at the job location.

Skills

- **Machine Learning & Predictive Modeling:** Scikit-learn, TensorFlow/Keras, Regression Analysis, Time Series Forecasting (ARIMA, SARIMA), Interpolation Techniques
- **Data Processing & Engineering:** Data Pipelines (Apache Airflow), Data Cleaning and Standardization (Databricks, Pandas, NumPy), Database Management (MySQL)
- **Large Language Models (LLMs) & Advanced Techniques:** LangChain, Retrieval-Augmented Generation (RAG), Hugging Face Transformers, Fine-tuning Transformers (BART)
- **Data Analysis & Visualization:** Tableau, Matplotlib, Seaborn, Plotly, Statistical Analysis
- **Programming Languages:** Python (Pandas, NumPy, SciPy, Scikit-learn), SQL
- **Cloud Platforms:** AWS (S3, Glue, Redshift, Kinesis, Lambda)
- **MLOps:** Docker, MLflow
- **Other Skills:** Git/GitHub, Feature Engineering, Hyperparameter Tuning, NLP

Experience

Consultant Data Scientist, Machine Learning Reply, Munich, Germany (Feb 2024 - June 2024, Full-time)

Provide advanced machine learning solutions for Stadtwerk am See:

- Built a scalable Q&A and classification system using Haystack, LangChain, and Streamlit, reducing response times by 25% through efficient knowledge base retrieval with Mistral and transformer-based pipelines.
- Improved answer retrieval accuracy by 20% by fine-tuning transformers like BART and using Retrieval-Augmented Generation (RAG) to deliver precise, context-aware responses.
- Achieved 90% precision in zero-shot classification of multilingual customer feedback using OCR-processed data and RoBERTa XLM, automating service request prioritization in real-time.
- Deployed the Streamlit app on AWS EC2 using Docker for containerization, ensuring scalability and real-time predictions.

Research Data Scientist, Salzburg Research m.b.H., Salzburg, Austria (July 2022 - Dec 2023, Full-time)

- Researched Energy Metrics for Edge-Cloud Systems: Defined crucial metrics to gauge energy consumption in distributed edge-cloud setups.
- Applied machine learning models on Amazon EC2 and Amazon S3 to analyze energy consumption patterns, identifying peak usage times and suggesting off-peak periods to users, resulting in a 12% reduction in overall energy consumption.

- Developed an LSTM model with an optimal window period to predict trends in energy consumption, improving resource allocation and operational efficiency by 15%.

- Utilized autoregressive models (e.g., ARIMA) for capturing long-term dependencies in energy data, enabling accurate forecasting and real-time adjustments in energy distribution.

- Performed feature engineering to identify key indicators affecting energy usage, improving model accuracy and enabling actionable insights for energy-saving strategies.

- Employed a decision tree-based ensemble method (e.g., Random Forest) to classify energy usage patterns, providing a clear understanding of consumption behavior and improving decision-making processes.

Data Analyst, LTM Paderborn, Paderborn, Germany (October 2021 - May 2022, Part-time)

- Collected and integrated data from MATLAB simulations, storing results in an Csv and Parquet.

- Conducted interpolation calculations using SciPy and developed regression models with scikit-learn to predict force based on displacement direction and other measurements, improving prediction accuracy by 20%.

- Designed interactive Tableau dashboards to visualize simulation results and predictive insights, reducing report generation time by 40%.

Analytics Engineer, Nextgen Healthcare, Bengaluru, India (June 2016 - July 2018, Full-time)

- Analyzed large-scale healthcare datasets, uncovering patient and operational trends to support clinical decision-making and enhance patient outcomes.

- Developed and automated data cleaning workflows using SQL and Python to preprocess healthcare data, ensuring accuracy and reliability for downstream analyses.

- Conducted exploratory data analysis (EDA) using Python (Pandas, NumPy) and statistical techniques to identify key patterns and anomalies in patient demographics, diagnoses, and treatment data.

- Built interactive dashboards and reports in Tableau to visualize patient treatment outcomes, disease prevalence trends, and resource utilization, improving insights for stakeholders.

- Performed root-cause analysis of data quality issues, proposing solutions to improve data collection processes and reduce errors by 20

Education

- **M.Sc. in Computer Science, Universität Paderborn, Germany**

    - Master Thesis: Logic analysis of Noctua cluster job logs using machine learning. Developed classification models (70% accuracy) to identify failing jobs and regression models (93% R^2 score) to predict resource needs (memory, CPU, time) for improved cluster utilization.

- **B.Eng. in Information Science, National Institute of Engineering, Mysuru, India**

    - Specialized in Software Engineering

Projects

- **Text Analysis for Argument Detection and Emotion Recognition**

    - Gathered data through web scraping (discussions, reviews). Built a Logistic Regression model to identify arguments and understand emotions within them.

    - Enhanced analysis with Bag-of-Words and Word2Vec for better text understanding. Reusable analysis pipelines created.

- **Smart Quality System (Product Fault Detection)**

    - Developed an automated monitoring system using historical data (vibration, logs) to identify potential printing quality issues.

    - Trained an AI classifier achieving 92% accuracy in detecting faults.